

Machine learning worksheet set-3

1. which of the following is an application of clustering?

Ans. Market trend prediction(B)

2. On which data type, we cannot perform cluster analysis?

Ans. None(D)

3. Netflix's movie recommendation system uses?

Ans. All the above(D)

4. The final output of Hierarchical clustering is?

Ans. The tree representing how close the datapoints are to each other(B)

5. Which of the step is not required in K-means clustering?

Ans. None(D)

6. Which is the following wrong?

Ans. K-nearest neighbor as same as k-means(c)

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

Ans. single link, complete link, Average link(D)

8. Which of the following are true?

Ans. Clustering analysis is negatively affected by heteroscedasticity(B)

9. In the figure above, if you draw a horizontal line on y-axis for $y=2$. What will be the number of clusters formed?

Ans. 2(A)

10. For which of the following tasks might clustering be a suitable approach?

Ans. . Given sales data from a large number of products in a supermarket, estimate future sales for each of these products(A)

11. Given, six points with the following attributes:

Ans. option(A)

12. Given, six points with the following attributes:

Ans.option(B)

13.What is the importance of clustering?

Clustering is important in data analysis and data mining applications. It is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups (clusters). Clustering can be done by the different no.

14.How can I improve my clustering performance?

Ans.There are several methods to improve the performance clustering algorithm

First of all try to compare with one that works well and compare the results.

Secondly,compare the time between the both algorithms,if u have set of good answers,then you can analyze how quality of the solution improves with the time

Thirdly,try your algorithms with several instances of the problem one not too challenging ,one medium and one hard and finally, using evolution to optimize clustering parameters of your clustering algorithms,